UC Davis
UC Davis Previously Published Works

Title

Elucidating the genetic basis of antioxidant status in lettuce (Lactuca sativa)

Permalink

Journal

ISSN

Authors

Damerum, Annabelle
Selmes, Stacey L
Biggi, Gaia F
et al.

Publication Date

DOI

Copyright Information

Peer reviewed

ARTICLE

Elucidating the genetic basis of antioxidant status in lettuce (*Lactuca sativa*)

Annabelle Damerum[1], Stacey L Selmes[1], Gaia F Biggi[1], Graham JJ Clarkson[1,2], Steve D Rothwell[2], Maria José Truco[3], Richard W Michelmore[3], Robert D Hancock[4], Connie Shellcock[4], Mark A Chapman[1] and Gail Taylor[1]

A diet rich in phytonutrients from fruit and vegetables has been acknowledged to afford protection against a range of human diseases, but many of the most popular vegetables are low in phytonutrients. Wild relatives of crops may contain allelic variation for genes determining the concentrations of these beneficial phytonutrients, and therefore understanding the genetic basis of this variation is important for breeding efforts to enhance nutritional quality. In this study, lettuce recombinant inbred lines, generated from a cross between wild and cultivated lettuce (*Lactuca serriola* and *Lactuca sativa*, respectively), were analysed for antioxidant (AO) potential and important phytonutrients including carotenoids, chlorophyll and phenolic compounds. When grown in two environments, 96 quantitative trait loci (QTL) were identified for these nutritional traits: 4 for AO potential, 2 for carotenoid content, 3 for total chlorophyll content and 87 for individual phenolic compounds (two per compound on average). Most often, the *L. serriola* alleles conferred an increase in total AOs and metabolites. Candidate genes underlying these QTL were identified by BLASTn searches; in several cases, these had functions suggesting involvement in phytonutrient biosynthetic pathways. Analysis of a QTL on linkage group 3, which accounted for >30% of the variation in AO potential, revealed several candidate genes encoding multiple MYB transcription factors which regulate flavonoid biosynthesis and flavanone 3-hydroxylase, an enzyme involved in the biosynthesis of the flavonoids quercetin and kaempferol, which are known to have powerful AO activity. Follow-up quantitative RT-PCR of these candidates revealed that 5 out of 10 genes investigated were significantly differentially expressed between the wild and cultivated parents, providing further evidence of their potential involvement in determining the contrasting phenotypes. These results offer exciting opportunities to improve the nutritional content and health benefits of lettuce through marker-assisted breeding.

INTRODUCTION

Dietary consumption of plant phytochemicals from fruits and vegetables has been linked to positive health effects[1–5] Antioxidants (AOs) may contribute to this benefit since they can scavenge free radicals, such as singlet oxygen and superoxide radicals, potentially reducing oxidative damage to cellular components,[6,7] although this role remains controversial and is not universally accepted.[8] These and other phytochemicals also have important roles in plants in defence against pests, pathogens, and UV light, attraction of pollinators and competitive interactions with other plants (reviewed in ref. 9). The largest group of phytonutrients is the phenolic compounds, including phenolic acids and flavonoids (anthocyanins, anthocyanidins, flavones, flavonols, flavanones, proanthocyanins and isoflavones[5]) and these have been shown to possess powerful AO activity *in vitro*.[10–13] Carotenoids; yellow, orange and red terpenoids, are another group of plant compounds with AO activity, acting as accessory pigments during photosynthesis to quench the excited state of chlorophyll, and also to provide colouration.[14,15] Many carotenoids have pro-vitamin A activity due to the presence of vitamin A as part of their structure, making them an important nutrient in the human diet, reducing the risk of respiratory diseases and blindness.[16] The value of the photosynthetic pigment chlorophyll as an important phytochemical in foods has been underappreciated. Chlorophyll derivatives extracted from spinach have

recently been shown to prevent DNA damage of human lymphocytes *in vitro* in a dose-dependent manner,[17] suggesting they are key contributors to the overall AO potential of foods.

Although lettuce is not usually acknowledged as being a rich source of beneficial phytochemicals, it does contain phenolic compounds, vitamins C and E, and carotenoids.[18,19] Lettuce consumption can improve cholesterol metabolism in rats and can stimulate the AO capacity of blood plasma in both humans and rats.[20,21] Beneficial phenolic compounds in lettuce include chicoric acid (also called dicaffeoyltartaric acid), chlorogenic acid (also known as caffeoyl quinic acid) and the flavonoid quercetin.[18,22,23] Chicoric acid extracted from lettuce inhibits both lipid peroxidation and cyclo-oxygenase enzyme activities; [23]chlorogenic acid is effective at inhibiting the hypermethylation of DNA, which is characteristic of tumour cells.[24] Finally, quercetin has potential anti-cancer properties, arresting A549 lung cancer cell lines *in vitro*.[25] Despite the effects of these phenolics in isolation, it has been suggested that eating whole foods rich in natural sources of these beneficial compounds is more effective than relying solely on dietary supplements.[26–28] Thus there is increasing pressure to develop new and novel germplasm with enhanced nutritional quality and to enable breeding programmes to use molecular markers effectively, with a clear understanding of the chemistry underlying nutritional traits.

Extensive DNA polymorphism data are available for lettuce. More than 35 000 lettuce genes have been analysed for single feature

[1]Centre for Biological Sciences, University of Southampton, Life Sciences, University Road, Southampton SO17 1BJ, UK; [2]Vitacress Limited, Lower Link Farm, St Mary Bourne, Andover, Hampshire SP11 6DB, UK; [3]The Genome Centre and the Department of Plant Sciences, University of California, Davis, CA 95616, USA and [4]The James Hutton Institute, Invergowrie, Dundee DD2 5DA, UK
Correspondence: G Taylor (g.taylor@soton.ac.uk)

polymorphisms (SFPs) using a high-density Affymetrix GeneChip microarray[29] and SFPs have been mapped to create an ultra-dense, gene-based, genetic linkage map for lettuce using a recombinant inbred line (RIL) population generated from a cross between *Lactuca sativa* (cultivated lettuce) and *L. serriola* (wild 'prickly' lettuce; http://chiplett.ucdavis.edu/).[30] This resource has been useful in determining the genetic basis for traits such as disease resistance and shelf-life in lettuce.[31–33] We therefore used this extensively characterized mapping population to investigate the genetic basis of AO potential. The aim of this study was to identify quantitative trait loci (QTL) determining AO potential, total carotenoid, chlorophyll and phenol content, and levels of individual metabolites, as well as to identify candidate genes underlying these QTL. This provides underpinning information to develop molecular tools for breeding lettuce with enhanced nutritional qualities. Previously, QTL have been identified for AO capacity in tomato fruits,[34] anthocyanin in raspberries,[35] carotenoid content in maize,[36] and chlorophyll and AO potential in lettuce,[37] but to our knowledge, none have linked these traits to underlying metabolic signatures and candidate genes.

MATERIALS AND METHODS

Plant material

RIL mapping population. Sixty F_9 RILs, generated from a cross between cultivated lettuce (Lactuca sativa cv. Salinas) and wild lettuce (*L. serriola* accession UC96US23), along with representatives of the parental lines, were used as the mapping population in this study. The RILs investigated are a subset of a total population of 213 RILs developed and characterized by the Compositae Genome Project (http://chiplett.ucdavis.edu/), which were determined to be highly informative during previous investigations.[32] Five *L. sativa* cultivars, denoted C1-C5, including two red types (C1; Daredevil and C2; SSC 3025) and three green (C3; Frontrunner, C4; Thriller, C5; Carlsbad) were obtained from Shamrock Seeds Company (UK).

Plant growth

Glasshouse experiments. Nine replicates of each of the RILs and parental lines were planted in a fully randomised blocked design (1–3 replicates per block and three blocks), with positions randomly selected using Minitab 14.0 (Minitab Inc., Philadelphia, PA, USA). Plants were grown in $70 \times 70 \times 80$ mm pots containing blended peat, seed and modular growing media, at pH 5.5 (Vapogro, Kekkilä and Avoncrop Ltd, Windsor, UK). Initially, four seeds were sown per pot and thinned following germination so that only one plant per pot remained to grow to maturity. Day temperatures ranged from 18 ℃ to 27 ℃ and night temperatures averaged 18 ℃, with day length approximately 16 h. Pots were watered from below when required. Following 5 weeks growth, whole plants at the rosette stage were harvested and leaves were ground to a fine powder in liquid N_2 before storage at –80 ℃.

Field experiment. Lettuce seeds were planted in September 2009 within a commercial crop of Lollo Rosso lettuce on a farm in Azenha do Mar, Odemira, Portugal (37°47′28″N, 8°79′18″W). Nine biological replicates were planted across a fully randomised and blocked design (see Glasshouse experiments), spaced at 0.1 m intervals in 1.2 m $\times$ 35.0 m beds and each block was surrounded with two rows of Cos lettuce to minimise edge effects.[32] Plants were irrigated from above when required and all other conditions were as those for the surrounding commercial crop. Whole plants were harvested at the rosette stage after 5 weeks of growth and transported under refrigeration to the University of Southampton, UK, where leaves were snap frozen and stored at –80 ℃.

Determination of AO potential

The Ferric Reducing Antioxidant Power assay (FRAP assay; refs.[38,39]) was used to estimate the total AO potential of the RILs and RIL parents grown in the glasshouse and controlled environment experiments, according to the method revised by Payne *et al.*[39] (See Supplementary Information for details).

Extraction and quantification of chlorophyll and carotenoid content

Three 1 cm diameter discs were taken from the fourth true leaf of each plant of the glasshouse grown RILs at point of harvest, avoiding major veins; one from the tip and one from either side of the mid-rib vein. Pigments were extracted from leaf discs by incubating in microfuge tubes containing 500 µl of dimethylformamide in the dark at 4 ℃ for >48 h. Absorbance of the extracts was measured at 647, 664 and 480 nm in a cuvette spectrophotometer (U-2000, Hitachi, Wokingham, UK). Chlorophyll a, b, total chlorophyll and carotenoid concentration (all µg/ml) were calculated.[40]

Determination of phenolic content

Phenols were extracted as outlined by Llorach *et al.*[19] with modifications. Briefly, leaf material ground in liquid N_2 was freeze-dried and 0.1–0.2 g of lyophilised leaf material was resuspended in 20 volumes of methanol:water:formic acid (25:24:3), vortexed rapidly and extracted in the dark at 4 ℃ for 30 min under continuous agitation. Samples were centrifuged (10 min, 13 000 rpm) and the supernatant was saved. The pellet was resuspended and re-extracted as described above, and the second supernatant was combined with the first. Extracts were kept in the dark at –20 ℃ until further analysis.

Total phenolic content. The enzymatic assay for total phenolic content was as outlined by Stevanto *et al.*,[41] with modifications. A total of 100 µl of the above supernatant was diluted 10-fold with water and added to 900 µl of reaction buffer (0.1 M potassium phosphate buffer (pH 8.0), containing 20 mM hydrogen peroxide, 30 mM 4-aminophenazone and 100 U/ml horseradish peroxidase). Following a 5-min reaction period absorbance of each sample in triplicate was measured at 500 nm in spectrophotometric cuvettes. Aqueous solutions of catechin (0.1–1 mM), previously utilised as a standard for measuring phenolic content in lettuce,[18] were utilised to generate a calibration curve from which total phenolic content of each sample was calculated as catechin equivalents, mg/ml dry weight (DW).

Identification and quantification of individual phenolics. A known concentration of the flavonol morin was added as an internal standard to the extracted phenols. *Liquid chromatography–mass spectrometry* (LC/MS)/MS was conducted on a Thermo HPLC system, consisting of an Acela autosampler and an Acela 600 pump (for further details see Supplementary Information). Compounds were identified using their UV absorption characteristics and parent and daughter ion masses as described.[19] Relative quantification was achieved from the parent ion peak area, corrected according to the peak area of the morin internal standard.

QTL analysis and identification of candidate genes

A dense linkage map of the RIL mapping population based on genic SFP markers was already available for QTL analysis (http://chiplett.ucdavis.edu/). A framework map consisting of 613 markers spaced approximately 3 cM apart across the 9 linkage groups (LGs) was used for the QTL analysis (Table 1; Supplementary Table S1). QTL mapping was conducted using composite interval mapping in Windows QTL Cartographer Ver. 2.5.[42] Chromosome walk speed was set at 1 cM and the logarithm of odds (LOD) threshold for declaring a significant QTL ($P < 0.05$) was estimated for each trait by permutation tests with 1000 iterations.[43] QTLs were plotted using MapChart 2.2.[44] Co-localising QTLs were defined as two or more QTL with overlapping LOD intervals. Candidate genes within major QTL were identified in BLASTn searches based on their similarity to genes annotated and reported in the literature as having roles which could influence the levels of secondary metabolites (see Supplementary Information). Single nucleotide polymorphisms (SNPs) were identified in the predicted coding regions of genes under the LG3 AO QTL by aligning sequencing reads of *L. serriola* (UC96US23) and 4 of the RILs determined to have within the top 10 AO potential and 4 of the RILs in the bottom 10 lowest AO potential (reads for all of the highest and lowest ranking RILs for AO potential were not available) to the *L. sativa* cv. Salinas reference genome sequence and where sequencing reads were available, identifying SNP haplotypes. cDNA sequences of candidate genes for *L. sativa* cv. Salinas and *L. serriola* were downloaded via GenBank, translated using the ExPASy tool[45] and aligned using Clustal[46] to identify non-synonymous amino acid substitutions and deletions.

Quantitative RT-PCR

Real-time qPCR was conducted to evaluate the expression of 10 candidate genes selected from the genomic regions underlying the LG3 AO QTL for the

wild and cultivated parents of the mapping population. Three biological replicates of each parent from the field trial were analysed in duplicate for the candidates, along with two reference genes *ACT* and *40S* (see Supplementary Information for details).

Statistical analysis

For the phenotype data, two sample *t*-testing and one-way analysis of variance (ANOVA) with post hoc Tukey's testing were conducted on raw phenotype data using Minitab 16 (Minitab Ltd.) and mean data were evaluated via Pearson's correlation coefficient analysis using SigmaPlot (Systat Software Inc.). Data were normalised by log-transformation when required. Differential expression between the wild and cultivated parents determined by quantitative RT-PCR was identified by two sample *t*-tests in R version 3.2.2.[47]

RESULTS

Phenotyping the RIL mapping population

AO potential. The AO potential of *L. serriola* acc. UC96US23, the wild parent of the RILs, was over threefold greater than that of *L. sativa* cv. Salinas, the cultivated parent (38.68 ± 7.72 vs. 9.83 ± 0.53 mmol, respectively; one-way ANOVA, $F_{3,32}=11.38$, $P < 0.001$). No significant differences were observed between the RIL with the highest AO potential (59.17 ± 11.69 mmol) and *L. serriola* nor the RIL with the lowest AO potential (11.19 ± 0.96 mmol) and *L. sativa*; however, there was evidence of transgressive segregation in that some RILs had a higher AO potential than the *L. serriola* parent (Figure 1A).

Total carotenoid and chlorophyll content. Both chlorophyll and carotenoid content were higher in the wild parent than in the cultivated one (one-way ANOVA, $F_{3,31}=58.63$, $P < 0.01$ and $F_{3,31}=27.09$, $P < 0.001$, respectively). Although carotenoid content was similar between *L. sativa* (20.52 ± 2.79 mg/m) and the lowest scoring RIL (22.30 ± 1.12 mg/m), the RIL scoring the greatest value for these traits (35.28 ± 1.20 mg/m) was significantly higher than *L. serriola* (29.69 ± 1.34 mg/m; Figure 1; $F_{3,31} = 27.09$, $P < 0.001$), again suggesting the presence of transgressive segregation. Similarly, chlorophyll content was significantly higher in the RIL with the highest value than *L. serriola* (22212 ± 16.69 vs. 184.07 ± 13.25 mg/m; Figure 1C; $F_{3,31} = 58.63$, $P < 0.01$).

Phenolic profile. Clear differences in total phenolic content were detected between wild and cultivated lettuce (38.76 ± 4.65 vs. 22.25 ± 1.25 catechin equivalent, mg/g dry weight, respectively; two-sample *t*-test, $t_{(8)}=-3.43$, $P \leqslant 0.01$). Individual phenolic compounds were further quantified by LC/MS/MS. Visual inspection of the LC-MS profiles revealed clear qualitative and quantitative differences in phenolic composition for the two parent lines (Figure 2). Metabolites such as caffeoyl tartaric acid (CTA), 5-p-coumaroylquinic acid (5-CoQA), caffeoyl quinic acid (CQA), di-pCT, kaempferol glucuronide (K-3Gc), quercetin-3-glucoronide (Q-3Gc), quercetin-3-malonylglucoside (Q-3MG) and kaempferol-3-malonylglucoside (K-3MG; peaks 1–3, 6 and 8–11, respectively, on Figure 2) were present at greater concentrations in *L. serriola* than *L. sativa*, whilst concentrations of di-CTA (DCTA; chicoric acid) and dicaffeoyl quinic acid (DCQA; peaks 5 and 7, respectively; Figure 2) were comparable and caffeoyl malic acid (peak 4; Figure 2) was the only phenolic compound noticeably present at greater abundance in *L. sativa*.

A total of 23 individual metabolites were detected from the leaf samples of the RILs grown in the glasshouse environment. These consisted of two CTA isoforms (CTA1 and CTA2), two di-caffeoyl tartaric acid isoforms (DCTA and mDCTA), caffeoyl malic acid (CMA), three caffeoyl quinic acid isoforms (CQA1, CQA2, CQA3), two di-caffeoylquinic acid isoforms (DCQA and 3,5-DCQA), two 5-coumaroylquinic acid isoforms (5-CoQA1 and 5-CoQA2), five quercetin derivatives (Q-3G, Q-3Gc, Q-3MG, Q-3MG-7Gc and Q-3MG-7G), two luteolin derivatives (L-7G and L-7Gc), two kaempferol derivatives (K-3Gc and K-3MG) and two unknown compounds, denoted
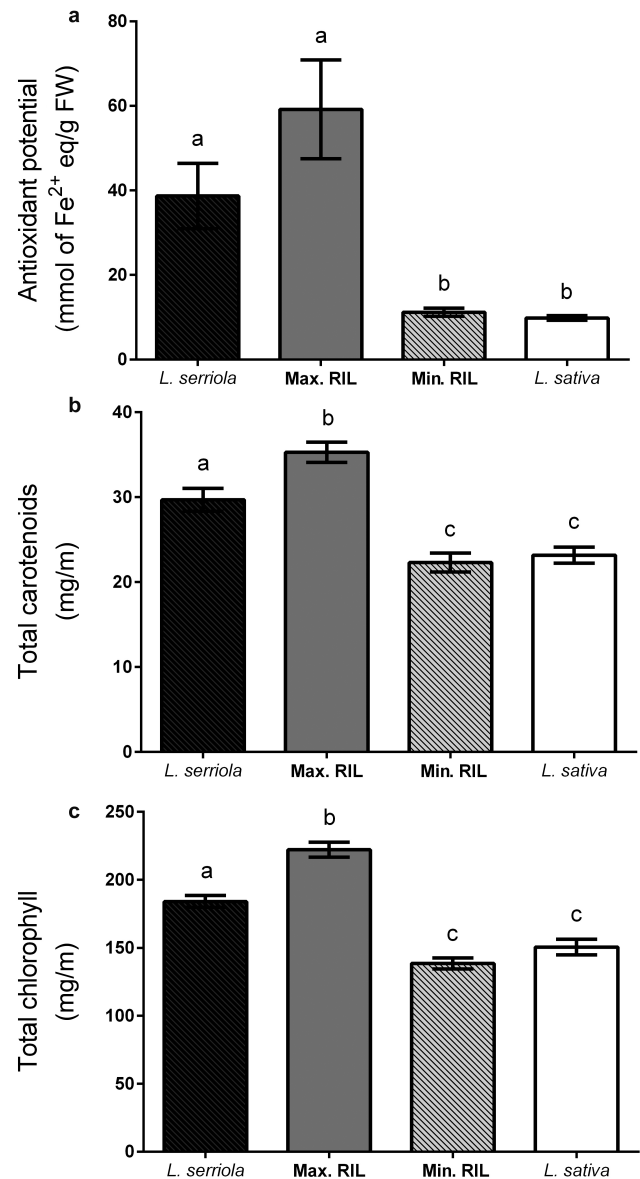
Figure 1. Phenotyping the RIL population. Analysis of antioxidant-related phenotypes in the parental lines (*L. sativa* and *L. serriola*), highest RIL (Max RIL) and lowest RIL (low Min RIL). Letters indicate significant differences. (**a**) total antioxidant potential (one-way ANOVA; $F_{3,32} = 11.38$, $P \leqslant 0.001$), (**b**) total carotenoids ($F_{3,31} = 27.09$, $P \leqslant 0.001$) and (**c**) total chlorophyll ($F_{3,31} = 58.63$, $P \leqslant 0.01$). Bars represent the mean $\pm$ standard error.

331 and 347 based on their m/z ratios (Supplementary Table S2). The above phenolics were also identified in the field grown RILs, excluding 5-CoQA2.

Relative quantities of each phenolic compound were estimated by comparison of the relative peak area to that of an internal standard. The top four most abundant metabolites were consistent amongst the field and glasshouse environments with chicoric acid (DCTA) found to be the most abundant in the RILs, comprising >30% of the total phenolic compounds in both the field and glasshouse grown RILs. Quercetin 3-malonylglucoside (Q-3MG) was identified as the second most abundant in both the field and glasshouse trials (18% and 16%, respectively). This was followed by CQA1 (11%) and Q-3Gc (10%) in the glasshouse grown RILs and Q-3Gc (10%) and CQA1 (9%) from the field.
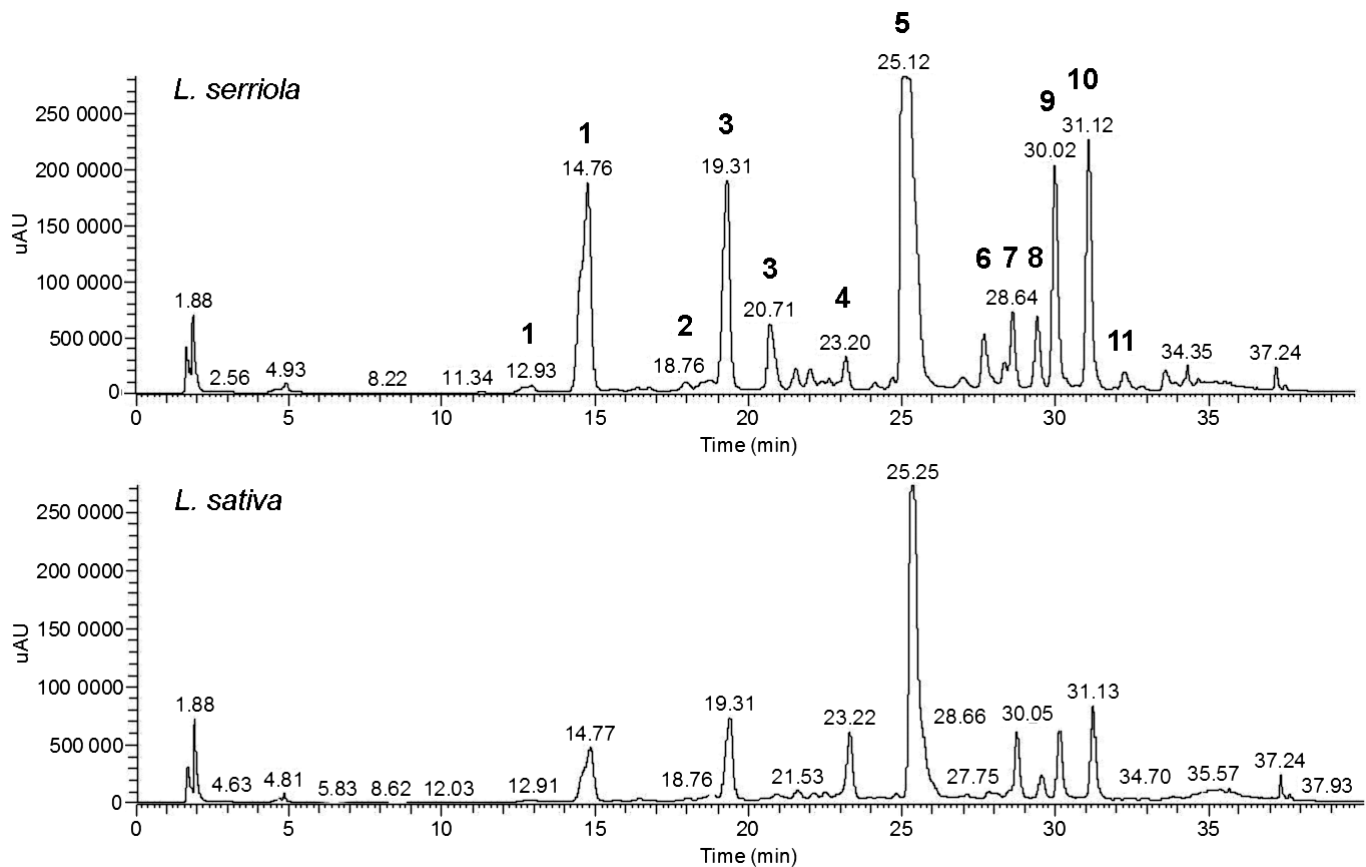
Figure 2. Comparison of the phenolic profiles of *L. serriola* and *L. sativa*. LC-MS profiles of the RIL parents, *Lactuca serriola* and *L. sativa*. Figure shows the relative maxima of absorbance of samples against retention time in (minutes), also displayed above each peak. Peaks: (1) CTA; (2) 5-CoQA (5-p-coumaroylquinic acid); (3) caffeoyl quinic acid; (4) caffeoyl malic acid; (5) di-CTA; (6) di-pCT; (7) dicaffeoyl quinic acid; (8) kaempferol glucuronide; (9) quercetin-3-glucuronide; (10) quercetin-3-malonylglucoside; (11) kaempferol-3-malonylglucoside. Multiple peaks for the same compound indicate isoforms.

Relative concentrations of the most abundant metabolites detected in lettuce were compared against *L. serriola* and *L. sativa* and for the top four RILs measured to have the highest (denoted HAO lines 1–4) and four RILs with the lowest (LAO lines 1–4) AO potential (Figure 3). DCTA concentration was significantly higher in the wild parent in comparison to the cultivated parent (one-way ANOVA, $F_{3,44} = 26.26$, $P < 0.001$) and this was also seen for CTA ($F_{3,44} = 26.26$, $P < 0.001$). Transgressive segregation of metabolites was often observed with concentrations usually higher in the HAO line than in the cultivated parent, seen for Q-3MG ($F_{3,44} = 6.43$, $P < 0.01$), CQA ($F_{3,44} = 12.34$, $P < 0.001$), Q-3G ($F_{3,44} = 13.97$, $P < 0.001$), CTA, DCQA ($F_{3,44} = 17.39$, $P < 0.001$), L-7G ($F_{3,44} = 10.31$, $P < 0.01$) and K-3MG ($F_{3,44} = 6.33$, $P < 0.01$), though there were no differences of relative CMA concentration amongst any of the lines ($F_{3,44} = 2.27$, $P > 0.05$). In most cases, the HAO line also had greater metabolite concentrations than the LAO, excluding for CMA. There were no differences in relative metabolite concentration between *L. sativa* and the LAO line excluding DCQA and no differences were observed between *L. serriola* and the HAO line.

AO potential is correlated with shelf life and various metabolites. Pearson's correlation coefficient analysis of mean trait data revealed several significant correlations. AO potential was found to be positively correlated with shelf life ($P < 0.05$, Table 1), measured in the same RIL subset by Zhang *et al.*[32] Total chlorophyll and carotenoids were strongly positively correlated with each other ($P < 0.001$) but negatively correlated with AO potential ($P < 0.01$).

Relative abundances of Q-3MG-7Gc, K-3MG, the CQA isoforms 2 and 3, and 5-CoQA2 were all found to be positively correlated with AO potential measured for the glasshouse grown RILs. Shelf life was positively correlated with the flavonoids Q-3MG-7Gc and K-3MG ($P < 0.05$) and negatively correlated with total carotenoids ($P < 0.05$). AO potential was not found to significantly correlate with total phenolics and 5-CoQA2 was the only metabolite to significantly positively correlate with phenolic content ($P < 0.01$).

Relationship between AO potential and total phenolics. AO potential and total phenolics of the four RILs measured to have the highest antioxidant potential (HAO lines) and the four with the lowest (LAO lines) were reassessed in a subsequent trial alongside five commercial varieties. Generally, AO potential increased with total phenolics, with the HAO and LAO lines typically clustering (Figure 4). The two red cultivars C1 and C2 showed superior AO and phenolic status, but all other cultivars (C3–C5) clustered with the majority of the LAO lines (Figure 4). The interaction between AO potential and phenolic content was not as expected for HAO2 and LAO4, with HAO2 clustering with the LAO lines 1–3 and LAO4 with the other HAO lines (Figure 4).

QTL analysis
A linkage map composed of 613 SFP markers distributed over the nine chromosomal LGs (http://chiplett.ucdavis.edu/) was utilised for QTL analysis. A total of 38 QTL from 24 traits were detected for the field trial and 62 QTL from 30 traits for the glasshouse trial, with QTL distributed across all nine LGs (Supplementary Table S5, Supplementary Figure S1).
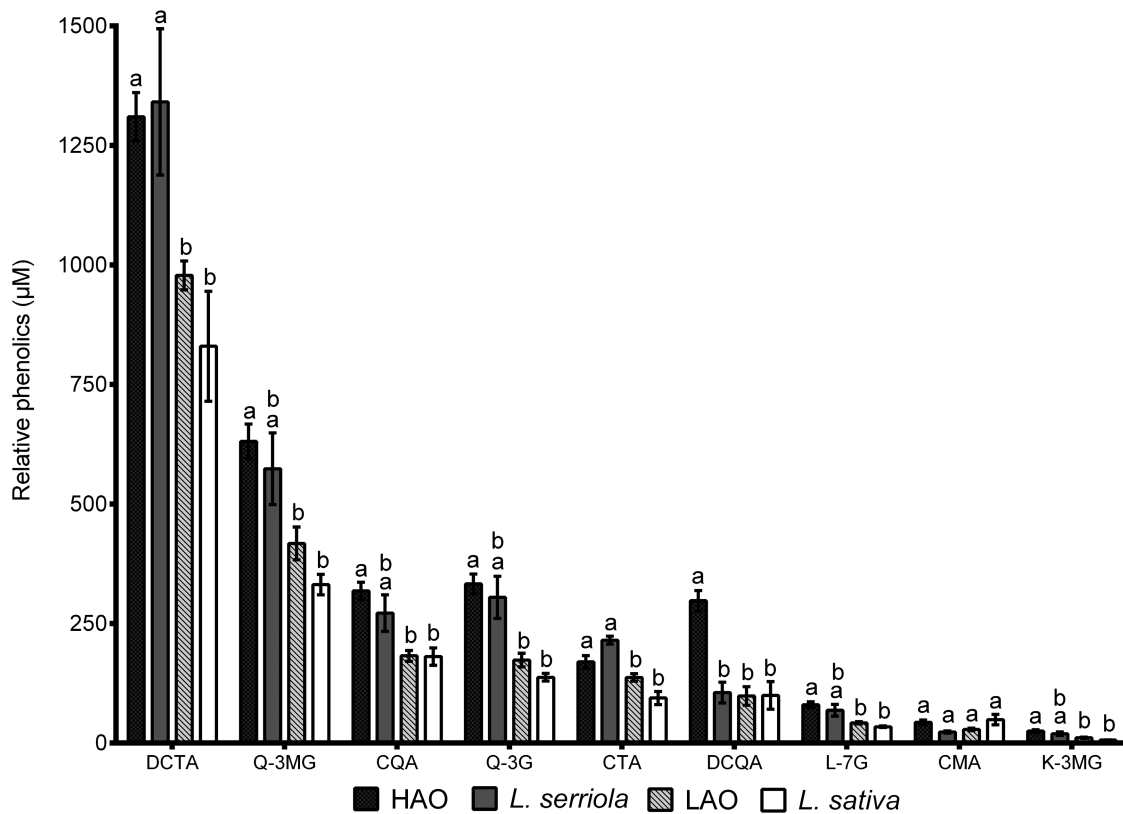
Figure 3. Comparison of relative phenolic concentrations amongst *L. serriola, L. sativa* and the high and low antioxidant RILs. Relative concentration of phenolics DCTA, Q-3MG, CQA, Q-3G, CTA, DCQA, L-7G, CMA and K-3MG for the parents and the four RILs measured to have the highest (HAO) and lowest (LAO) antioxidant status. Bars represent the mean ± standard error, with letters indicating significant differences (one-way ANOVA; see text for details).

AO potential. Four significant QTL were detected for AO potential measured in the glasshouse grown RILs, on LGs 3, 4, 7 and 9, accounting for 30%, 12%, 16% and 9% of the phenotypic variation (PV), respectively. For the QTL on LGs 3, 4 and 7 the alleles inherited from *L. serriola* resulted in an increase in the trait value whereas for the fourth QTL allele inheritance from *L. sativa* caused an increase in the trait.

Total carotenoid and chlorophyll content. Two QTL for total chlorophyll were detected on LGs 3 and 7, with allele inheritance from *L. sativa* and *L. serriola*, respectively, explaining 25% and 16% of the PV. Both of these QTL co-located with those for AO potential from

FRAP. An additional QTL for chlorophyll a was detected on LG9, explaining 12% of the variation for this trait and with allele inheritance from *L. sativa*, but in contrast to the previous QTL, this did not co-locate with AO potential.

Phenolic compounds. In both the field and glasshouse studies a single QTL for total phenolic content was identified. Each explained over 30% of the PV but were found on different LGs (8 and 4, respectively). QTLs were identified for 18 out of the 23 metabolites detected in the field grown RILs (excluding CTA1, CQA1, unknown metabolite 347, 5-CoQA1 and 5-CoQA2) and explained between 10% and 33% of the PV. QTLs were detected

Table 1. Pearson's correlation coefficient analysis of traits measured from the RIL population.

Trait	Shelf life	AO	Phenolics	CHL	CAR	Q-3MG-7Gc	K-3MG	CQA2	CQA3	5-CoQA2
Shelf life	1									
AO	0.304*	1								
Phenolics	−0.115	0.199	1							
CHL	−0.254	−0.414**	−0.298	1						
CAR	−0.275*	−0.402**	−0.277	0.987	1					
Q-3MG-7Gc	0.417*	0.434*	0.309	−0.181	−0.181	1				
K-3MG	0.433*	0.409*	0.042	0	0.002	0.648***	1			
CQA2	0.184	0.461**	0.157	−0.062	−0.062	0.505**	0.209	1		
CQA3	0.264	0.439*	0.354	−0.14	−0.14	0.634***	0.396*	0.911***	1	
5-CoQA2	0.005	0.384*	0.471**	−0.162	−0.165	0.504**	0.475**	0.416*	0.49**	1

Correlation of mean trait values measured in the glasshouse trial, with shelf life data taken from previous investigations on the same RIL subset, taken from Zhang *et al*. (ref. 32). ***$P < 0.001$; **$P < 0.01$ and *$P < 0.05$ levels. AO, antioxidant potential; CHL, total chlorophyll content; CAR, total carotenoid content. Correlations for all traits measured can be seen in Supplementary Table S3.
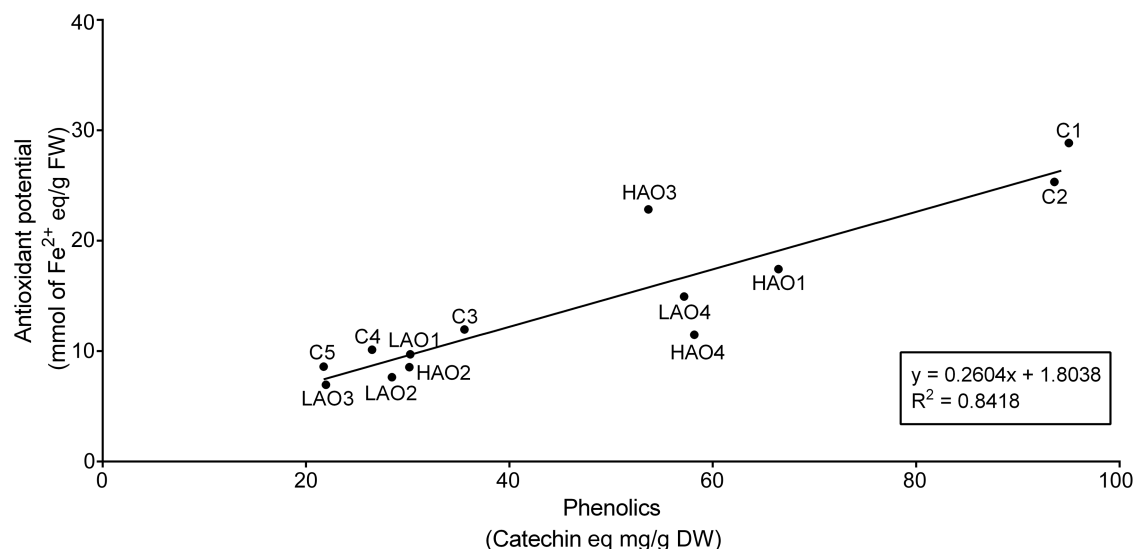
Figure 4. Relationship between antioxidant potential and total phenolics. Linear regression between antioxidant potential and total phenolics of the four RILs determined to have the highest AO potential (HAO 1–4) and four with the lowest AO potential (LAO 1–4), along with five *L. sativa* cultivars (C 1–5).

for 19 of the 23 phenolic compounds measured in the glasshouse grown RILs (with the exception of Q-3MG-7Gc, the two uncharacterised metabolites and 5-CoQA1). Individual QTL explained between 8% and 47% of the PV.

QTL hotspots. Almost every LG harboured regions of co-locating QTL (where LOD intervals overlapped), here defined as 'QTL hotspots', some of which were for the same trait measured in the two different growing environments (Supplementary Table S5). QTL for four phenolic compounds, DCTA, CTA1 and CTA2 in the glasshouse grown RILs and K-3MG in the field grown RILs were found to co-locate on LG1. QTL for the quercetin metabolite Q-3Gc on LG2 was found to co-locate with 3,5-DCQA and mDCTA from the glasshouse grown RILs. A similar cluster of QTL was identified on LG5, with a Q-3G measured from the glasshouse grown RILs co-locating with CMA and CTA1 measured from the field grown RILs. QTLs for CTA1 measured from the field and glasshouse trials were found to co-locate on LG5 and LG6 and a QTL for total phenolics measured from the field trial co-located with 5-CoQA.

LG3 contained the highest number of QTL (19 detected) and also a QTL hotspot of several individual compounds, which mapped to the centre of the LG within a 13 cM range. This included two kaempferol derivatives, from the glasshouse trial (K-3Gc) and the field trial (K-3MG), and a quercetin derivative from the glasshouse trial (Q-3MG). This region also corresponded to the large effect AO potential QTL plus QTL for total chlorophyll and carotenoids (Supplementary Table S5). This suggests that this region of the genome may be worthy of further investigation and development of molecular markers for breeding.

Each of the four QTL detected for AO potential measured in the glasshouse grown RILs were found to co-locate with other traits. In two instances, QTL for AO potential (on LG3 and LG7) were found to co-locate with QTLs for total chlorophyll and carotenoids (Supplementary Table S5). On LG4, the QTL for AO potential co-located with another for kaempferol derivative K-3MG also measured from the glasshouse grown RILs. On LG3 and LG9, QTL for AO potential co-located with QTL for Q-3MG, from the field and glasshouse trials, respectively. A QTL for Q-3MG-7Gc measured from the glasshouse RILs was also found to co-localise with the QTL for AO potential on LG9 (Supplementary Table S5).

Identification of candidate genes

For the large effect AO QTL on LG3, candidate genes in the genomic region were identified in the lettuce reference genome sequence and their putative functions inferred, based on sequence similarity to the annotated genomes of *Arabidopsis thaliana* and *Solanum lycopersicum*. A total of 285 genes were identified from approximately 50 Mbp of the genome corresponding to the QTL region. Several genes acting in the phenylpropanoid pathway and known to directly influence secondary metabolism were identified within this QTL (Table 2). A gene which acts in the flavonoid biosynthetic pathway, flavanone 3-hydroxylase (F3H), was found within this region, which encodes a key enzyme in the synthesis of flavonoids quercetin and kaempferol.[48]

Two other enzymes acting in the phenylpropanoid pathway were identified (Figure 5); caffeoyl-CoA O-methyltransferase (CCoAOMT) which is involved in lignin biosynthesis[49] and ferulate-5-hydroxylase (F5H), which is also involved in lignin biosynthesis but has recently been implicated in inducing the biosynthesis of anthocyanins under photooxidative stress in Arabidopsis.[50] Two MYB transcription factors, one of which is production of anthocyanin pigment 2 (PAP2) and a closely related R2R3 class MYB transcription factor MYB114, both known to regulate the conversion of flavonol precursors (dihydrokaempferol and dihydroquercetin) to anthocyanin precursors (anthocyanidins) in the flavonoid biosynthetic pathway, were also located in this region.[51,52] (Table 2, Figure 6) PAP2 was positioned within the estimated AO QTL peak and three distinct copies of MYB114, each spaced >20 kbp apart were in this region.

Genes encoding zeaxanthin epoxidase and geranylgeranyl pyrophosphate (GGPP) synthase, two genes involved in carotenoid biosynthesis,[53,54] were also detected in this region (Table 2). Other notable candidates include a gene encoding ascorbate peroxidase (APX), involved in reactive oxygen species metabolism,[55] a xyloglucan endotransglucosylase/hydrolase (XTH) involved in cell wall modification,[56] and another MYB transcription factor (*MYB44*), found to delay leaf senescence when overexpressed in *A. thaliana*.[57] Candidate gene analysis therefore revealed several genes which warrant further functional investigation.

For each of the 285 candidate genes identified within the AO QTL on LG3, inheritance of SNPs in coding regions for a selection of 8 RILs measured to have amongst the highest and lowest AO potential was determined by aligning genomic reads of *L. serriola* (UC96US23) with those of the RILs (Figure 6). SNP haplotype of

Table 2. Candidate genes identified within the AO QTL on LG3.

A. thaliana ccession	Lactuca accession[a]	Functional description	Reference
At1g66390	Letassy_X1_8017	Production of anthocyanin pigment 2 (PAP2) protein	51
	Serrassy_T_P2_17469	R2R3 class of MYB transcription factors, involved in anthocyanin biosynthesis	
AT1G66380	Letassy_X1_6767	MYB114 R2R3 class of MYB transcription factors closely related to the production of anthocyanin pigment type MYBs involved in anthocyanin biosynthesis	50
	Serrassy_T_P2_17850		
At3g51240	Letassy_X1_4796	Flavanone 3-hydroxylase, enzyme involved in flavonoid biosynthesis	47
	Serrassy_T_P2_12444		
At4g36220	Letassy_X1_2126	Ferulate-5-hydroxylase, enzyme involved in lignin and anthocyanin biosynthesis	49
	Serrassy_T_P2_6166		
AT4G34050	Letassy_X1_23118	Caffeoyl-CoA O-methyltransferase, enzyme involved in lignin biosynthesis	48
	Serrassy_T_P2_7154		
AT4G36810	Letassy_X1_21865	GGPP synthase, enzyme involved in terpenoid (includes carotenoids) biosynthesis	52
	Serrassy_T_P2_6921		
At5g67030	Letassy_X1_1094	Zeaxanthin epoxidase precursor, enzyme involved in zeaxanthin (common carotenoid) biosynthesis	53
	Serrassy_T_P2_6429		
AT5G65730	Letassy_X1_23118	XTH 7, regulated by MYB41 and has a putative role in cell expansion	55
	Serrassy_T_P2_7154		
AT5G67300	Letassy_X1_21649	MYB44 R2R3 class of MYB transcription factor found to have multiple roles in ABA signalling and leaf senescence	56
	Serrassy_T_P2_5693		
AT4G35000	Letassy_X1_9546	Ascorbate peroxidase, enzyme involved in reactive oxygen species metabolism in leaf peroxisomes	54
	Serrassy_T_P2_46272		

[a] Both *L. sativa* and *L. serriola* accessions tabulated.

the RILs was generally as expected, with the HAO RILs inheriting the wild parent SNP allele and LAO RILs having the cultivated parent allele and this was particularly pronounced in the region corresponding to the peak of QTL, which contained *PAP2* and *MYB114*.

Analysis of candidate genes

For the 10 candidates selected (Table 2), relative gene expression for both cultivated and wild parents was determined by qRT-PCR, in an attempt to identify differential expression. Five out of the 10 candidate genes were found to be differentially expressed between the parents (Figure 7), including *PAP2* (A, $P < 0.05$), *MYB114* (B, $P < 0.05$), *F3H* (C, $P < 0.01$), *F5H* (D, $P < 0.05$) and *GGPS* (E, $P < 0.01$) and are proposed as the best candidate genes from the 10 originally selected. With the exception of *MYB114*, all were more highly expressed in the wild parent. Three of these genes, *PAP2*, *MYB114* and *GGPS*, were located within the estimated QTL peak.

Comparison of the *L. sativa* cv. Salinas and *L. serriola* cDNA sequences revealed a number of non-synonymous amino acid changes in *MYB114*, including a seven amino acid deletion in the cultivated parent protein sequence (A, Supplementary Figure S3). The cultivated F3H protein had one amino acid difference from valine to isoleucine in comparison to the wild protein (B, Supplementary Figure S3) and F5H had three non-synonymous differences (C, Supplementary Figure S3). The *L. sativa* APX protein had four amino acid differences from *L. serriola* (D, Supplementary Figure S3).

DISCUSSION

Understanding AO potential in lettuce

Our investigations of AO potential of a lettuce RIL population showed that phenolics, carotenoids and chlorophyll were important contributors to this quantitative trait, with evidence of transgressive segregation, perhaps indicating the complementary action of alleles inherited from both parents.[58] Such lines offer exciting prospects for the development of lettuce with enhanced nutritional value. Transgressive segregants for fresh weight of tomatoes,[59] milling quality in rice,[60] aluminium tolerance in sorghum[61] and grain yield in durum wheat[62] have been proposed as potential sources for the improvement of these quantitative traits.

The phenolic composition of both wild and cultivated lettuce contrasted both qualitatively and quantitatively, with relative abundances differing amongst the RIL parents and with wild lettuce containing higher overall concentration of phenolic compounds. The greatest differences were observed for derivatives of CTA, caffeoyl quinic acid (CQA), quercetin-3-glucuronide (Q-3GC) and quercetin-3-malonylglucoside (Q-3MG; peaks 1, 3, 9 and 10; Figure 2), which were present in trace amounts or at greatly reduced levels in the cultivated parent. These dramatic differences are likely to significantly compromise the nutritional quality of the cultivated lettuce in comparison to its wild counterpart, subsequently impacting on associated health benefits following consumption. Metabolites such as CQA are lost or broken down when cooked,[63] making lettuce an important source of these phenolics in the diet given that it is consumed raw and so even slight changes in metabolite abundance will have a major impact on health. DCTA(chicoric acid) was the most abundant phenolic in the RILs, consistent with other investigations measuring the phenolic composition of lettuce cultivars,[18,19] with lettuce recognised as being one of the main European dietary sources of chicoric acid.[63] Kaempferol derivatives were the least abundant phenolics detected, which are usually measured in trace amounts in comparison to other flavonols such as quercetin,[64] a derivative of which (Q-3MG) was the second most abundant phenolic in both the field and glasshouse grown RILs.

Differences in phenolic concentrations were also observed amongst the RILs according to AO status. Increased concentrations of the most abundant phenolics such as chichoric acid, quercetin derivatives Q-3MG and Q-3G, CQA, CTA, DCQA, L-7G and K-3MG were observed in the HAO lines, though out of these metabolites only CQA and K-3MG were found to be significantly positively correlated with AO potential. The lack of significant correlation between AO potential and total phenolics (Table 1), despite the positive relationship observed for the extreme HAO and LAO RILs (Figure 4), is likely to reflect the wide genetic background of the RIL population.

QTL for AO potential co-locate with numerous metabolites

For the first time to our knowledge, we have linked genomic regions in lettuce underlying AO status to candidate genes, using genomic resources developed for lettuce.[30] In the present invest-
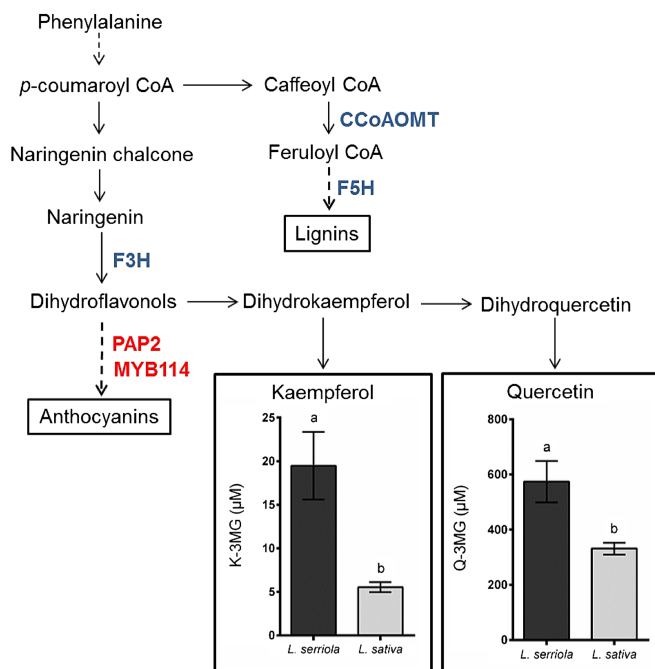
Figure 5. Candidates genes involved in the phenylpropanoid pathway. A summary of the phenylpropanoid pathway detailing roles of the candidate genes identified within the LG3 QTL region. Genes encoding three enzymes were identified: F3H, flavanone 3-hydroxylase; CCoAOMT, caffeoyl-CoA O-methyltransferase; F5H, ferulate-5-hydroxylase, coloured in blue and two MYB family transcription factors: PAP2; PRODUCTION OF ANTHOCYANIN PIGMENT 2 and MYB114, coloured in red. Differences in concentration of K-3MG and Q-3MG in wild and cultivated lettuce are shown. Letters indicate where significant differences were observed, as measured by two-sample *t*-test ($t_5 = 4.03$, $P \leqslant 0.01$, and $t_5 = 3.58$, $P \leqslant 0.05$, respectively).

igation, two QTL for total carotenoid content were identified on LG3 and LG7, which based on current literature are the only QTL so far determined for carotenoid content in a leafy vegetable. Although QTL for chlorophyll have previously been determined to vary depending on growing environment,[37] QTL for chlorophyll content on LG3, 7 and 9 measured from the glasshouse trial confirm those previously identified from a UK field trial,[32] providing strong evidence for consistency in these QTL. The four QTL identified for AO potential mapped to LG3, 4, 7 and 9, with the largest effect QTL on LG3 (LOD score 8.7) accounting for almost one-third (30.2%) of the PV for this trait, thus a large-effect QTL. Alleles inherited from *L. serriola* increased AO potential for all QTLs excluding alleles at the QTL on LG7, which was inherited from *L. sativa* in the majority of RILs with a higher AO potential. This was to be expected given that the wild parent was measured to possess an overall greater AO potential than cultivated lettuce.

It is perhaps unsurprising that QTL for total chlorophyll and carotenoids were found to co-locate on LG3 and 7, given their coordinated synthesis and intimate relationship in the chloroplasts as part of photosynthetic complexes[65,66] and as the biosynthetic pathways are commonly linked through the precursor GGPP.[67] GGPP was one of the candidates identified in the LG3 hotspot region (Table 2) which was found to be more highly expressed in the wild parent and could explain the co-location of QTL for chlorophyll and carotenoids observed. The co-locating QTL on LG7 for total chlorophyll and carotenoids explained 16–18% of the PV and possession of the *L. serriola* allele was found to increase trait values, which was expected as the wild parent had significantly increased concentrations of both pigments in comparison with *L. sativa*. QTL for total

chlorophyll and carotenoids on LG3 had effects in the opposite direction as would be predicted with respect to the phenotype and measured gene expression, with the *L. sativa* allele increasing trait value. *Trans* arrangement of positive alleles has previously been linked with transgressive segregation of traits from an interspecific cross of tomato;[59] a phenomenon which was observed for both chlorophyll and carotenoids in the present investigation of lettuce. Co-location of QTL for total chlorophyll and carotenoids on LG3 and LG7 with total AO potential supports the findings of Hayashi *et al*.[37] Despite this, although chlorophyll and carotenoid were positively correlated with each other, they were measured to be negatively correlated with AO potential in the present investigation. The large effect QTL for AO on LG3 (30.2% variation explained) also co-located with a QTL for Q-3MG, which explained 17.8% of the PV. QTL for AO on LG9 (16.2 % variation explained) also co-located with Q-3MG (16.8% variation explained), as well as the quercetin metabolite Q-3MG-7G (14.8% variation explained). As quercetin metabolites act as powerful AOs,[68] it is therefore likely that fluctuations can notably affect total AO potential, suggesting that we have identified an important metabolic trait underpinning AO potential in this lettuce mapping population.

On each LG there was evidence of co-location of QTL for metabolites and in some cases, QTL for the same trait mapped to the same position in both environments. For example, QTL consistent across environments included CTA (CTA1) and on another LG, QTL for CTA1 and chicoric acid (DCTA) measured from the field trial co-located with a QTL for the kaempferol derivative K-3MG (kaempferol-3-malonylglucoside) measured from the glasshouse (Supplementary Table S5). Total phenolics measured from the field grown RILs co-located with 5-CoQA2 (5-*p*-coumaryl quinic acid isomer 2) measured from the glasshouse, which has a key role in the phenylpropanoid pathway for secondary metabolism biosynthesis.[69] It has been known for some time that genes with a related function often cluster into operons in bacteria and there is growing evidence for the clustering of genes encoding secondary metabolites in plants.[70] For example, metabolic gene clusters for terpenoid biosynthesis have now been found in oat and *Arabidopsis* and more recently in the wild legume *Lotus japonicas*,[71] which may explain the many instances of QTL for different metabolites co-locating to the same region. QTL which have a consistent effect across different growing environments are considered more stable, thus are valued for use in breeding,[72] but this was not observed for all traits. Given that only a subset of the total RIL population was used for this study, the ability to detect small effect QTL was likely to be limited as population size has been demonstrated to limit the sensitivity of QTL detection.[73] Another possible explanation is that significant genotype × environment (G × E) interactions are occurring, which is perhaps unsurprising given that environmental factors are known to have an impact on secondary metabolism.[74] Indeed, significant G × E interactions have been detected for AO and CHL QTL;[37] however, similar analyses of data collected from the glasshouse and field environments in present investigations would not be appropriate due to differences in experimental design.

Interestingly, a cluster of QTL co-located to the centre of LG3 for dry weight following nutrient limitation and drought recovery using this population,[75] indicating a potential link between abiotic stress and AO potential, though direct comparisons of QTL were not possible due to differences in linkage maps utilised. Future work to analyse phenotype data using compatible mapping resources could reveal co-locations of QTL for abiotic stress tolerance and nutritional quality to the same genomic position, highlighting a strong target for marker-assisted breeding.

Identifying candidate genes for AO potential
Using the Lettuce Genome Resource, along with the previously sequenced genomes of *A. thaliana* and *S. lycopersicum*, several
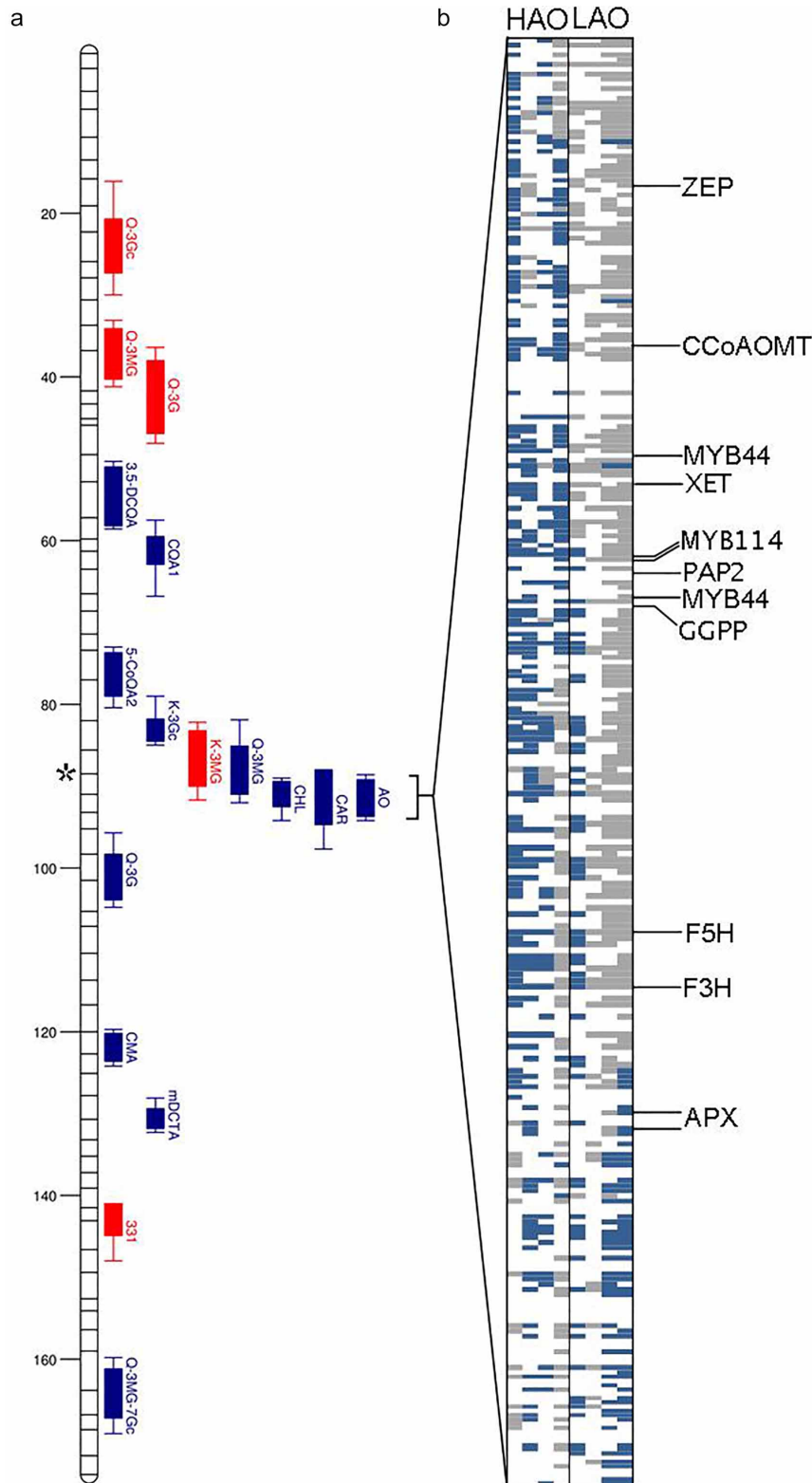
Figure 6. SNP genotyping of gene candidate under the LG 3 QTL or the AO QTL hotspot on LG 3 (**a**) SNP allele inheritance of the selected high antioxidant (HAO) RILs (columns 1–4) and low antioxidant (LAO) RILs (columns 5–8) was determined (**b**). Wild parent SNP allele inheritance is denoted in blue, inheritance of the cultivated parent allele is in grey and white boxes show where there is either missing sequencing data or no target SNPs present. Each row indicates one of the 285 candidate genes identified within the QTL region, highlighting the positions of the candidates described in Table 2. Position of the QTL hotspot is denoted by *with the scale bar to the left of the LG image showing genetic distance in centimorgans. Filled bars indicate the one-LOD interval and the line extensions show the two-LOD interval, for QTL identified in the field trial (blue) and glasshouse trial (red).
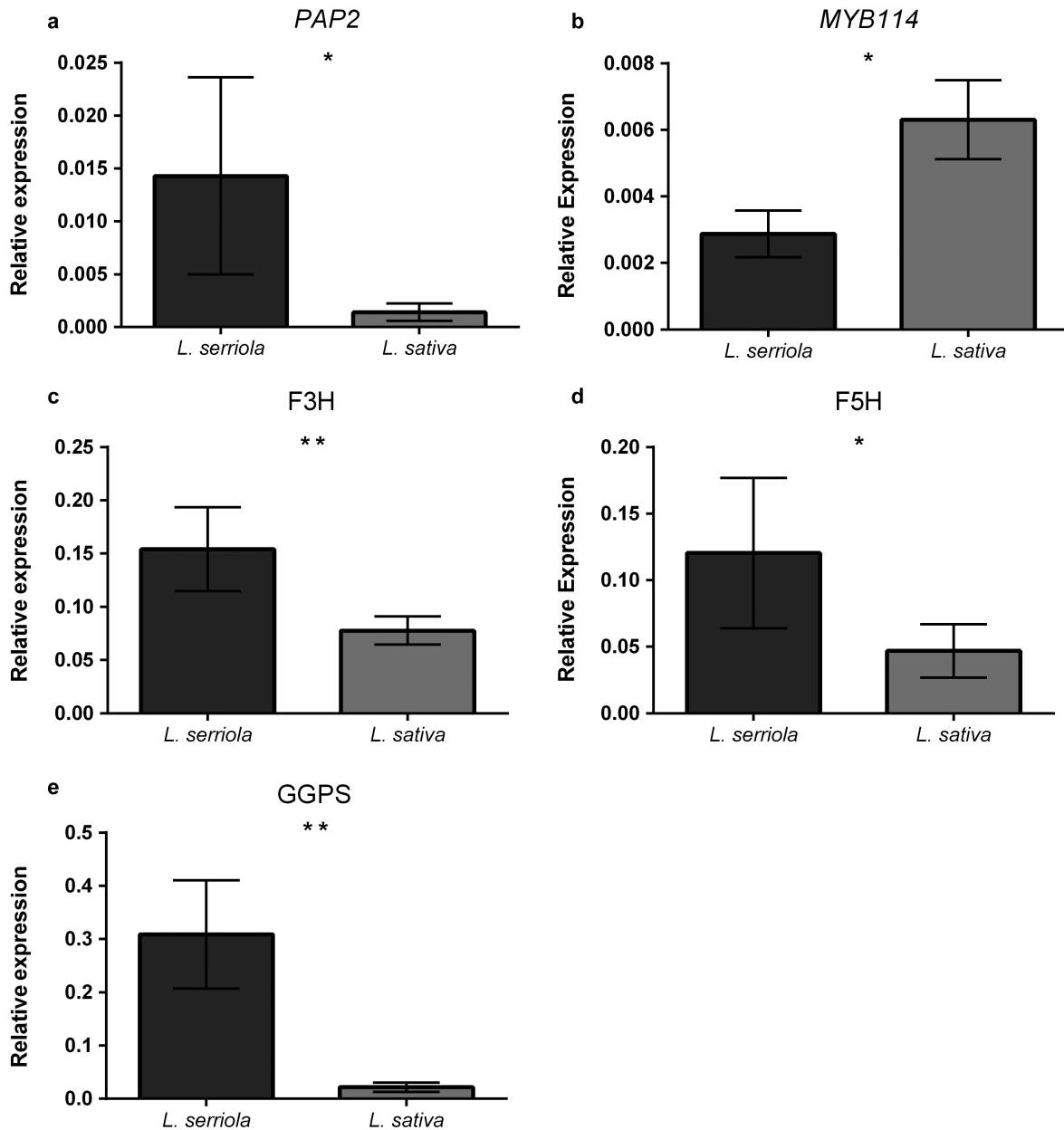
Figure 7. Comparison of candidate gene expression in wild and cultivated lettuce. Quantitative RT-PCR expression analysis of gene candidates detected within the AO QTL on LG3. PAP2 (**a**), MYB114 (**b**), F3H (**c**), F5H (**d**) and GGPS (**e**) were determined to be differentially expressed between wild (*L. serriola*) and cultivated (*L. sativa*) lettuce, with level of significance denoted as *$P < 0.05$, **$P < 0.01$. Bars represent mean ± standard error. Five remaining candidates which were not differentially expressed can be viewed in Supplementary Figure S2.

promising gene candidates explaining variation in AO potential in the lettuce RILs were identified on LG3, including two MYB transcription factors (*PAP2* and *MYB114*) thought to regulate anthocyanin biosynthesis (Table 2). Expression analysis revealed that both *PAP2* and *MYB114* genes were differentially expressed (Figure 7), with expression increasing and decreasing, in the wild and cultivated parents, respectively. Anthocyanins are a subclass of flavonoids synthesised from dihydroflavonols known to be one of the major compounds controlling plant colour, particularly fruits and this is largely regulated by the MYB transcription factors.[51] The presence of anthocyanins in the red lettuce cultivars investigated (C1 and C2; Figure 4) is likely to have resulted in a higher AO and phenolic content compared to the green cultivars and the high AO

RILs, the latter of which are green-leaved, with no anthocyanin metabolites detected in this population. Mulabagal *et al.*[23] investigated the phenolic contents of red and green lettuce types and although one anthocyanin was identified in red lettuce (cyanidin-3-*O*-(6''-malonyl-b-glucopyranoside)), none were detected in green types, which is also supported by phenolic composition analysis by Llorach *et al.*[19] Enhancing the expression of *PAP1* and the highly similar *PAP2* (93% identity of the R2R3 domain) using activation-tagging in *Arabidopsis* resulted in increased expression of phenyl-propanoid biosynthesis genes, including phenylalanine ammonia lyase, the enzyme which initiates the phenylpropanoid pathway and chalcone synthase, the first enzyme acting in flavonoid biosynthesis,[76] which could explain how the increased expression of PAP2

observed could contribute to AO status, despite the lack of anthocyanins detected. MYB114 also has a role in regulating anthocyanin biosynthesis that is similar to PAP2, through interaction with basic helix-loop-helix (bHLH) proteins, in a mechanism which is highly conserved throughout the plant kingdom.[77] Given that effects of *MYB114* overexpression are dependent on overexpression of a corresponding bHLH transcription factor, this could explain how *L. sativa* had reduced phenolic content and AO status compared to *L. serriola*, despite exhibiting increased expression of *MYB114*. Synchronised increases in the expression of MYB and bHLH transcription factors may result in the red leaf phenotype observed in commercial lettuce types.

Another promising gene candidate identified within this region was *F3H*, which was more highly expressed in wild relative to cultivated lettuce. F3H is involved in the conversion of naringenin to the dihydroflavonols dihydrokaempferol and dihydroquercetin (Figure 5), which are the precursors for kaempferol and quercetin, respectively.[48] Both kaempferol and quercetin were present in higher concentrations in the wild parent than the cultivated parent (Figure 5), likely to result from an increased abundance of dihydroflavnols caused by increased levels of F3H and consistent with the former having a greater AO potential. Dihydroquercetin is essential not only as a precursor for quercetin metabolites, but also for flavonoids such as catechin and the proanthocyadins,[78] which may also contribute to AO potential. Derivatives of quercetin (Q-3MG-7Gc) and kaempferol (K-3MG) were found to be strongly positively correlated in the present investigation (Table 1), indicating tightly coordinated regulation of the biosynthesis of these flavonoids.

Ferulate-5-hydroxylase (F5H), an enzyme acting in the phenylpropanoid biosynthesis pathway,[79] was also found to be more abundant in wild lettuce. Knocking out *F5H* in *Arabidopsis* has revealed a range of phenotypes, affecting lignin biosynthesis, UV protection and response to wounding.[80] *F5H* mutants had increased expression of *MYB4*,[80] a negative regulator of chalcone synthase,[81] thus reducing flavonoid biosynthesis which is consistent with the reduced levels of flavonoids detected in cultivated lettuce in the present investigation.

Sustainable intensification and breeding for increased AO potential
Enhanced food security requires that we achieve 'more from less' and that yield enhancements in future crops must be complemented by higher nutritional value (Agri-tech Strategy, www.gov.co.uk). Many crop-breeding programmes are now dedicated to developing enhanced crop nutrition where wild progenitors of crops may be exploited for higher concentrations of target phytonutrients relative to those observed in their commercial counterparts.[82,83] Indeed, such an approach has been successfully deployed for many food crops including tomatoes,[34] berries[84,85] carrots[86] and potatoes.[87] This can be a powerful approach – broccoli florets from cultivated varieties were found to have between 3 and 10 μmol/g of health-benefitting glucosinolates, whilst wild species can contain 50–100 μmol/g.[83]

Here, the AO potential of cultivated lettuce (*L. sativa* cv. Salinas) was significantly lower than that of the wild progenitor (*L. serriola*), with notable differences in phenolic composition. Past artificial selection of lettuce for improved yield traits is likely to be linked to indirect selection against characteristics such as AO status since phenolics are known to have a bitter taste.[88] During evolution we have learned to reject bitter tastes and with >50 bitter taste receptors characterised, aversion is likely to have been crucial to survival.[88] A RIL (HAO3) which had comparable levels of AOs to the red varieties but a reduced phenolic content was identified in the present investigation (Figure 4), which could be utilised in future breeding programmes.

Co-incidentally, improving leaf nutritional quality may also afford greater plant protection from pests and diseases, given that many secondary metabolites have roles in defence against herbivore and pathogen attack[89] with mechanical wounding resulting in the accumulation of phenolic compounds in lettuce.[90] Microbial spoilage in particular has been shown to reduce the shelf life of lettuce.[91] Tomatoes genetically engineered to overexpress anthocyanins had an increased shelf life in comparison to wild-type-cultivated tomatoes, with the transgenic tomatoes demonstrating reduced susceptibility to the fungal pathogen *Botrytis cinerea*.[92] Leaf AO potential and derivatives of the phenolics, quercetin and kaempferol, measured in present investigations were found to positively correlate with shelf life, measured on the same RIL subset by Zhang and colleagues,[32] further supporting a link between phytochemicals and shelf life. A gene encoding a XTH involved in cell wall loosening and located within the QTL hotspot in LG3 has also been linked with shelf life, with the down-regulation of a XTH previously shown to increase the shelf life of lettuce,[93] though the XTH identified was not determined to be differentially expressed between the wild and cultivated parents in present investigations (Supplementary Figure S2). Clearly further investigations are required to dissect the link between nutritional quality and shelf life in lettuce.

Improving the phytonutrient content of widely consumed yet relatively nutritionally poor vegetables, such as lettuce, is an important target for plant breeding and here we identified several candidates controlling flavonoid biosynthesis within the large-effect QTL for AO potential; a number of which were shown to be differentially expressed between wild and cultivated lettuce. The QTL region underpinning these traits is a strong target for future breeding and on-going research is focusing on introgressing this genomic region into commercial lettuce breeding lines. At the same time, further proof of functionality through genome editing and other molecular routes is also underway. Taken together, this study provides the first detailed insight into lettuce phytonutrient traits and how they may be deployed in the future for an enhanced food plant, consumed widely and of global significance.

COMPETING INTERESTS

ACKNOWLEDGEMENTS

REFERENCES

1 Okabe S, Suganuma M, Hayashi M, Sueoka E, Komori A, Fujiki H. Mechanisms of growth inhibition of human lung cancer cell line, PC-9, by tea polyphenols. *Jpn J Cancer Res* 1997; **88**: 639–643.

2 Chatzi L, Apostolaki G, Bibakis I *et al*. Protective effect of fruits, vegetables and the Mediterranean diet on asthma and allergies among children in Crete. *Thorax* 2007; **62**: 677–683.

3 Moon H-S, Chung C-S, Lee H-G, Kim T-G, Choi Y-J, Cho C-S. Inhibitory effect of (-)-epigallocatechin-3-gallate on lipid accumulation of 3T3-L1 cells. *Obesity* 2007; **15**: 2571–2582.

4 Ueda M, Nishiumi S, Nagayasu H, Fukuda I, Yoshida K, Ashida H. Epigallocatechin gallate promotes GLUT4 translocation in skeletal muscle. *Biochem Biophys Res Commun* 2008; **377**: 286–90.

5 El Gharras H. Polyphenols: food sources, properties and applications – a review. *Int J Food Sci Technol* 2009; **44**: 2512–2518.

6 Halliwell B. Are polyphenols antioxidants or pro-oxidants? What do we learn from cell culture and *in vivo* studies? *Arch Biochem Biophys* 2008; **476**: 107–112.

7 Maiani G, Castón MJP, Catasta G *et al*. Carotenoids: actual knowledge on food sources, intakes, stability and bioavailability and their protective role in humans. *Mol Nutr Food Res* 2009; **53**: S194–218.

8 Martin C, Butelli E, Petroni K, Tonelli C. How can research on plants contribute to promoting human health? *Plant Cell* 2011; **23**: 1685–1699.

9 Hounsome N, Hounsome B, Tomos D, Edwards-Jones G. Plant metabolites and nutritional quality of vegetables. *J Food Sci* 2008; **73**: R48–65.

10 Kähkönen MP, Hopia AI, Vuorela HJ *et al*. Antioxidant activity of plant extracts containing phenolic compounds. *J Agric Food Chem* 1999; **47**: 3954–3962.

11 Duthie GG, Duthie SJ, Kyle JA. Plant polyphenols in cancer and heart disease: implications as nutritional antioxidants. *Nutr Res Rev* 2000; **13**: 79–106.

12 Stevenson DE, Hurst RD. Polyphenolic phytochemicals – just antioxidants or much more? *Cell Mol Life Sci* 2007; **64**: 2900–2916.

13 Song W, Derito CM, Liu MK, He X, Dong M, Liu RH. Cellular antioxidant activity of common vegetables. *J Agric Food Chem* 2010; **58**: 6621–6629.

14 Bartley GE, Scolnik PA. Plant carotenoids: pigments for photoprotection, visual attraction, and human health. *Plant Cell* 1995; **7**: 1027–1038.

15 Demmig-Adams B, Adams III WW. Antioxidants in photosynthesis and human nutrition. *Science* 2002; **298**: 2149–2153.

16 Ye X, Al-Babili S, Klöti A *et al*. Engineering the provitamin A (β-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science* 2000; **287**: 303–305.

17 Hsu C, Chao P, Hu S, Yang C. The antioxidant and free radical scavenging activities of chlorophylls and pheophytins. *Food Nutr Sci* 2013; **4**: 1–8.

18 Nicolle C, Carnat A, Fraisse D *et al*. Characterisation and variation of antioxidant micronutrients in lettuce (Lactuca sativa folium). *J Sci Food Agric* 2004; **84**: 2061–2069.

19 Llorach R, Martínez-Sánchez A, Tomás-Barberán FA, Gil MI, Ferreres F. Characterisation of polyphenols and antioxidant properties of five lettuce varieties and escarole. *Food Chem* 2008; **108**: 1028–1038.

20 Serafini M, Bugianesi R, Salucci M, Azzini E, Raguzzini A, Maiani G. Effect of acute ingestion of fresh and stored lettuce (*Lactuca sativa*) on plasma total antioxidant capacity and antioxidant levels in human subjects. *Br J Nutr* 2002; **88**: 615–623.

21 Nicolle C, Cardinault N, Gueux E *et al*. Health effect of vegetable-based diet: lettuce consumption improves cholesterol metabolism and antioxidant status in the rat. *Clin Nutr* 2004; **23**: 605–614.

22 Ferreres F, Gil I, Castan M, Toma FA. Phenolic metabolites in red pigmented lettuce (*Lactuca sativa*). Changes with minimal processing and cold storage. *J Agric Food Chem* 1997; **45**: 4249–4254.

23 Mulabagal V, Ngouajio M, Nair A, Zhang Y, Gottumukkala AL, Nair MG. *In vitro* evaluation of red and green lettuce (*Lactuca sativa*) for functional food properties. *Food Chem* 2010; **118**: 300–306.

24 Lee WJ, Zhu BT. Inhibition of DNA methylation by caffeic acid and chlorogenic acid, two common catechol-containing coffee polyphenols. *Carcinogenesis* 2006; **27**: 269–277.

25 Lu J, Papp LV, Fang J, Rodriguez-Nieto S, Zhivotovsky B, Holmgren A. Inhibition of mammalian thioredoxin reductase by some flavonoids: implications for myricetin and quercetin anticancer activity. *Cancer Res* 2006; **66**: 4410–4418.

26 Rowland I. Optimal nutrition: fibre and phytochemicals. *Proc Nutr Soc* 1999; **58**: 415–419.

27 Møller P, Loft S. Interventions with antioxidants and nutrients in relation to oxidative DNA damage and repair. *Mutat Res* 2004; **551**: 79–89.

28 Altunkaya A, Gökmen V. Effect of various anti-browning agents on phenolic compounds profile of fresh lettuce (*L. sativa*). *Food Chem* 2009; **117**: 122–126.

29 Stoffel K, van Leeuwen H, Kozik A *et al*. Development and application of a 6.5 million feature Affymetrix Genechip® for massively parallel discovery of single position polymorphisms in lettuce (*Lactuca* spp.). *BMC Genomics* 2012; **13**: 185.

30 Truco MJ, Ashrafi H, Kozik A *et al*. An ultra high-density, transcript-based, genetic map of lettuce. *G3 Genes | Genomes | Genet* 2013; **3**: 617–631.

31 Truco MJ, Antonise R, Lavelle D *et al*. A high-density, integrated genetic linkage map of lettuce (*Lactuca* spp.). *Theor Appl Genet* 2007; **115**: 735–746.

32 Zhang FZ, Wagstaff C, Rae AM *et al*. QTLs for shelf life in lettuce co-locate with those for leaf biophysical properties but not with those for leaf developmental traits. *J Exp Bot* 2007; **58**: 1433–1449.

33 McHale LK, Truco MJ, Kozik A *et al*. The genomic architecture of disease resistance in lettuce. *Theor Appl Genet* 2009; **118:** 565–580.

34 Rousseaux MC, Jones CM, Adams D, Chetelat R, Bennett A, Powell A. QTL analysis of fruit antioxidants in tomato using Lycopersicon pennellii introgression lines. *Theor Appl Genet* 2005; **111**: 1396–1408.

35 Kassim A, Poette J, Paterson A *et al*. Environmental and seasonal influences on red raspberry anthocyanin antioxidant contents and identification of quantitative traits loci (QTL). *Mol Nutr Food Res* 2009; **53:** 625–634.

36 Wong JC, Lambert RJ, Wurtzel ET, Rocheford TR. QTL and candidate genes phytoene synthase and zeta-carotene desaturase associated with the accumulation of carotenoids in maize. *Theor Appl Genet* 2004; **108**: 349–359.

37 Hayashi E, You Y, Lewis R, Calderon MC, Wan G, Still DW. Mapping QTL, epistasis and genotype × environment interaction of antioxidant activity, chlorophyll content and head formation in domesticated lettuce (*Lactuca sativa*). *Theor Appl Genet* 2012; **124**: 1487–1502.

38 Benzie IFF, Strain JJ. The Ferric Reducing Ability of Plasma (FRAP) as a measure of 'antioxidant power': the FRAP assay. *Anal Biochem* 1996; **239**: 70–76.

39 Payne AC, Mazzer A, Clarkson GJJ, Taylor G. Antioxidant assays – consistent findings from FRAP and ORAC reveal a negative impact of organic cultivation on antioxidant potential in spinach but not watercress or rocket leaves. *Food Sci Nutr* 2013; **1**: 439–444.

40 Wellburn AR. The spectral determination of chlorophyll a and chlorophyll b, as well as total carotenoids, using various solvents with spectrophotometers of different resolution. *J Plant Physiol* 1994; **144**: 307–313.

41 Stevanto R, Fabris S, Momo F. New enzymatic method for the determination of total phenolic content in tea and wine. *J Agric Food* 2004; **52**: 6287–6293.

42 Wang S, Basten CJ, Zeng Z-B. *Windows QTL Cartographer 2.5*. Department of Statistics, North Carolina State University, Raleigh, NC. 2012. http://statgen.ncsu.edu/qtlcart/WQTLCart.htm

43 Churchill GA, Doerge RW. Empirical threshold values for quantitative trait mapping. *Genetics* 1994; **138**: 963–971.

44 Voorrips RE. MapChart: software for the graphical presentation of linkage maps and QTLs. *J Hered* 2002; **93**: 77–78.

45 Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel R, Bairoch A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 2003; **31**: 3784–3788.

46 Sievers F, WilmA, Dineen D *et al*. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 2014; **7**: 539–539.

47 R Core Development Team. R: *A language and environment for statistical computing*. Vienna, Austria, 2008.

48 Owens DK, Crosby KC, Runac J, Howard BA, Winkel BSJ. Biochemical and genetic characterization of Arabidopsis flavanone 3 β-hydroxylase. *Plant Physiol Biochem* 2008; **46**: 833–843.

49 Ageorges A, Fernandez L, Vialet S, Merdinoglu D, Terrier N, Romieu C. Four specific isogenes of the anthocyanin metabolic pathway are systematically co-expressed with the red colour of grape berries. *Plant Sci* 2006; **170**: 372–383.

50 Maruta T, Noshi M, Nakamura M *et al*. Ferulic acid 5-hydroxylase 1 is essential for expression of anthocyanin biosynthesis-associated genes and anthocyanin accumulation under photooxidative stress in *Arabidopsis*. *Plant Sci* 2014; **219–220**: 61–68.

51 Allan AC, Hellens RP, Laing WA. MYB transcription factors that colour our fruit. *Trends Plant Sci* 2008; **13**: 99–102.

52 Zuluaga DL, Gonzali S, Loreti E *et al*. *Arabidopsis thaliana* MYB75/PAP1 transcription factor induces anthocyanin production in transgenic tomato plants. *Funct Plant Biol* 2008; **35**: 606–618.

53 Burke CC, Wildung MR, Croteau R. Geranyl diphosphate synthase: cloning, expression, and characterization of his prenyltransferase as a heterodimer. *Proc Natl Acad Sci* 1999; **96**: 13062–13067.

54 Baroli I, Do AD, Yamane T, Niyogi KK. Zeaxanthin accumulation in the absence of a functional xanthophyll cycle protects *Chlamydomonas reinhardtii* from photooxidative stress. *Plant Cell* 2003; **15**: 992–1008.

55 Shulaev V, Schlauch K, Mittler R. Cytosolic ascorbate peroxidase 1 is a central component of the reactive oxygen gene network of *Arabidopsis*. *Plant Cell* 2005; **17**: 268–281.

56 Cominelli E, Sala T, Calvi D, Gusmaroli G, Tonelli C. Over-expression of the *Arabidopsis* AtMYB41 gene alters cell expansion and leaf surface permeability. *Plant J* 2008; **53**: 53–64.

57 Jaradat MR, Feurtado JA, Huang D, Lu Y, Cutler AJ. Multiple roles of the transcription factor AtMYBR1/AtMYB44 in ABA signaling, stress responses, and leaf senescence. *BMC Plant Biol* 2013; **13**: 192.

58 Rieseberg LH, Archer MA, Wayne RK. Transgressive segregation, adaptation and speciation. *Heredity* 1999; **83**: 363–372.

59 DeVicente MC, Tanksley S. QTL analysis of transgressive segregation in an interspecific tomato cross. *Genetics* 1993; **134**: 585–596.

60 Dong Y, Tsuzuki E, Lin D *et al*. Molecular genetic mapping of quantitative trait loci for milling quality in rice (*Oryza sativa* L.). *J Cereal Sci* 2004; **40**: 109–114.

61 Caniato FF, Guimaraes CT, Schaffert RE *et al*. Genetic diversity for aluminum tolerance in sorghum. *Theor Appl Genet* 2007; **114**: 863–876.

62 Singh AK, Knox RE, Clarke JM *et al*. Genetics of pre-harvest sprouting resistance in a cross of Canadian adapted durum wheat genotypes. *Mol Breed* 2014; **33**: 919–929.

63 Clifford MN. Chlorogenic acids and other cinnamates – nature, occurrence, dietary burden, absorption and metabolism. *J Sci Food Agric* 2000; **80**: 1033–1043.

64 Hertog MGL, Hollman PCH, Venema DP. Optimization of a quantitative HPLC determination of potentially anticarcinogenic flavonoids in vegetables and fruits. *J Agric Food Chem* 1992; **40**: 1591–1598.

65 Vishnevetsky M, Ovadis M, Vainstein A. Carotenoid sequestration in plants: the role of carotenoid-associated proteins. *Trends Plant Perspect* 1999; **4**: 232–235.

66 Meier S, Tzfadia O, Vallabhaneni R, Gehring C, Wurtzel ET. A transcriptional analysis of carotenoid, chlorophyll and plastidial isoprenoid biosynthesis genes during development and osmotic stress responses in *Arabidopsis thaliana*. *BMC Syst Biol* 2011; **5**: 77–95.

67 Qin G, Gu H, Ma L *et al*. Disruption of phytoene desaturase gene results in albino and dwarf phenotypes in *Arabidopsis* by impairing chlorophyll, carotenoid, and gibberellin biosynthesis. *Cell Res* 2007; **17**: 471–482.

68 Dueñas M, Surco-Laos F, González-Manzano S, González-Paramás AM, Santos-Buelga C. Antioxidant properties of major metabolites of quercetin. *Eur Food Res Technol* 2011; **232**: 103–111.

69 Fraser CM, Chapple C. The phenylpropanoid pathway in *Arabidopsis*. *Arab B* 2011; **9**: e0152.

70 Chu HY, Wegel E, Osbourn A. From hormones to secondary metabolism : the emergence of metabolic gene clusters in plants. *Plant J* 2011; **66**: 66–79.

71 Krokida A, Delis C, Geisler K *et al*. A metabolic gene cluster in *Lotus japonicus* discloses novel enzyme functions and products in triterpene biosynthesis. *New Phytol* 2013; **200**: 675–690.

72 El-Soda M, Malosetti M, Zwaan BJ, Koornneef M, Aarts MGM. Genotype × environment interaction QTL mapping in plants: lessons from *Arabidopsis*. *Trends Plant Sci* 2014; **19**: 390–398.

73 Li X, Quigg RJ, Zhou J *et al*. A critical evaluation of the effect of population size and phenotypic measurement on QTL detection and localization using a large F2 murine mapping population. *Genet Mol Biol* 2006; **173**: 166–173.

74 Ramakrishna A, Ravishankar GA. Influence of abiotic stress signals on secondary metabolites in plants. *Plant Signal Behav* 2011; **6**: 1720–1731.

75 Hartman Y, Hooftman DAP, Uwimana B *et al*. Abiotic stress QTL in lettuce crop-wild hybrids: comparing greenhouse and field experiments. *Ecol Evol* 2014; **4**: 2395–2409.

76 Borevitz JO, Xia Y, Blount J, Dixon RA, Lamb C. Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *Plant Cell* 2000; **12**: 2383–2394.

77 Zimmermann IM, Heim MA, Weisshaar B, Uhrig JF. Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like BHLH proteins. *Plant J* 2004; **40**: 22–34.

78 Dixon R, Xie D, Sharma S. Proanthocyanidins – a final frontier in flavonoid research? *New Phytol* 2004; **165**: 9–28.

79 Maruta T, Noshi M, Nakamura M *et al*. Ferulic acid 5-hydroxylase 1 is essential for expression of anthocyanin biosynthesis-associated genes and anthocyanin accumulation under photooxidative stress in *Arabidopsis*. *Plant Sci*. 2014; **219–220**: 61–68.

80 Huang J, Bhinu V-S, Li X, Dallal Bashi Z, Zhou R, Hannoufa A. Pleiotropic changes in *Arabidopsis* F5H and sct mutants revealed by large-scale gene expression and metabolite analysis. *Planta* 2009; **230**: 1057–1069.

81 Zhao J, Zhang W, Zhao Y *et al*. SAD2, an importin-like protein, is required for UV-B response in *Arabidopsis* by mediating MYB4 nuclear trafficking. *Plant Cell* 2007; **19**: 3805–3818.

82 McCouch S, Baute GJ, Bradeen J *et al*. Agriculture: feeding the future. *Nature* 2013; **499**: 23–24.

83 Mithen R, Faulkner K, Magrath R, Rose P, Williamson G, Marquez J. Development of isothiocyanate-enriched broccoli, and its enhanced ability to induce phase 2 detoxification enzymes in mammalian cells. *Theor Appl Genet* 2003; **106**: 727–734.

84 Wang S, Lewers K. Antioxidant capacity and flavonoid content in wild strawberries. *J Am Soc Hortic Sci* 2007; **132**: 629–637.

85 Milivojevic J, Maksimovic V, Nikolic M, Bogdanovic J, Maletic R, Milatovic D. Chemical and antioxidant properties of cultivated and wild Fragaria and Rubus berries. *J Food Qual* 2011; **34**: 1–9.

86 Grassmann J, Schnitzler WH, Habegger R. Evaluation of different coloured carrot cultivars on antioxidative capacity based on their carotenoid and phenolic contents. *Int J Food Sci Nutr* 2007; **58**: 603–611.

87 Brown CR, Vales I, Yilma S *et al*. 'AmaRosa,' a red skinned, red fleshed fingerling with high phytonutrient value. *Am J Potato Res* 2012; **89**: 249–254.

88 Drewnowski A, Gomez-Carneros C. Bitter taste, phytonutrients, and the consumer: a review. *Am J Clin Nutr* 2000; **72**: 1424–1435.

89 Bennett RN, Wallsgrove RM. Secondary metabolites in plant defence mechanisms. *New Phytol* 1994; **127**: 617–633.

90 Kang H-M, Saltveit ME. Antioxidant capacity of lettuce leaf tissue increases after wounding. *J Agric Food Chem* 2002; **50**: 7536–7541.

91 Allende A, Aguayo E, Artés F. Microbial and sensory quality of commercial fresh processed red lettuce throughout the production chain and shelf life. *Int J Food Microbiol* 2004; **91**: 109–117.

92 Martin C, Bassolino L, Zhang Y, Schoonbeek H, Kiferle C, Perata P. Accumulation of anthocyanins in tomato skin extends shelf life. *New Phytol* 2013; **200**: 650–655.

93 Wagstaff C, Clarkson GJJ, Rothwell SD, Page A, Taylor G, Dixon MS. Characterisation of cell death in bagged baby salad leaves. *Postharvest Biol Technol* 2007; **46**: 150–159.

94 Hall T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 1999; **41**: 95–98.

95 Matvienko M, Kozik A, Froenicke L *et al*. Consequences of normalizing transcriptomic and genomic libraries of plant genomes using a duplex-specific nuclease and tetramethylammonium chloride. *PLoS One* 2013; **8**: e55913.

96 Chang S, Puryear J, Cairney J. A simple and efficient method for isolating RNA from pine trees. *Plant Mol Biol Rep* 1993; **11**: 113–116.

97 Untergasser A, Cutcutache I, Koressaar T *et al*. Primer3-new capabilities and interfaces. *Nucleic Acids Res* 2012; **40**: 1–12.

98 Borowski JM, Galli V, da Silva Messias R *et al*. Selection of candidate reference genes for real-time PCR studies in lettuce under abiotic stresses. *Planta* 2014; **239**: 1187–1200.

99 Andersen CL, Jensen JL, Ørntoft TF. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res* 2004; **64**: 5245–5250.

100 Ramakers C, Ruijter JM, Lekanne Deprez RH, Moorman AFM. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci Lett* 2003; **339**: 62–66.

101 Bernth Jensen JM, Petersen MS, Stegger M, Østergaard LJ, Møller BK. Real-time relative qPCR without reference to control samples and estimation of run-specific PCR parameters from run-internal mini-standard curves. *PLoS One* 2010; **5**: e11723.

102 Vandesompele J, De Preter K, Pattyn F *et al*. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 2002; **3**: Research0034.1–0034.11.